

<b>REPORT DOCUMENTATION PAGE</b>			Form Approved OMB NO. 0704-0188		
<p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p> <p>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</p>					
1. REPORT DATE (DD-MM-YYYY) 06-05-2015		2. REPORT TYPE Final Report		3. DATES COVERED (From - To) 22-Jul-2013 - 21-Jul-2014	
4. TITLE AND SUBTITLE Final Report: Workshop on Sensing and Analysis of High-Dimensional Data			5a. CONTRACT NUMBER W911NF-13-1-0273		
			5b. GRANT NUMBER		
			5c. PROGRAM ELEMENT NUMBER 611102		
6. AUTHORS Lawrence Carin			5d. PROJECT NUMBER		
			5e. TASK NUMBER		
			5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAMES AND ADDRESSES Duke University 2200 W Main Street Suite 710 Erwin Square Durham, NC 27705 -4677			8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211			10. SPONSOR/MONITOR'S ACRONYM(S) ARO		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S) 64542-CS-CF.1		
12. DISTRIBUTION AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited					
13. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.					
14. ABSTRACT This workshop brought together leading experts in the new field of analysis and acquisition of high-dimensional data, including compressive sensing (CS). The meeting focused on new theory, algorithms and application. In addition to having talks from many of the leading researchers from academia, there were talks from the members of the US government and industry, on directions for this new technology, including future research directions and programs.					
15. SUBJECT TERMS Sensing, workshop, analysis					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Lawrence Carin
a. REPORT UU	b. ABSTRACT UU	c. THIS PAGE UU			19b. TELEPHONE NUMBER 919-660-5270



## Report Title

Final Report: Workshop on Sensing and Analysis of High-Dimensional Data

### ABSTRACT

This workshop brought together leading experts in the new field of analysis and acquisition of high-dimensional data, including compressive sensing (CS). The meeting focused on new theory, algorithms and application. In addition to having talks from many of the leading researchers from academia, there were talks from the members of the US government and industry, on directions for this new technology, including future research directions and programs.

---

**Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:**

**(a) Papers published in peer-reviewed journals (N/A for none)**

Received

Paper

**TOTAL:**

**Number of Papers published in peer-reviewed journals:**

---

**(b) Papers published in non-peer-reviewed journals (N/A for none)**

Received

Paper

**TOTAL:**

**Number of Papers published in non peer-reviewed journals:**

---

**(c) Presentations**

Number of Presentations: 0.00

---

**Non Peer-Reviewed Conference Proceeding publications (other than abstracts):**

Received      Paper

**TOTAL:**

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

---

**Peer-Reviewed Conference Proceeding publications (other than abstracts):**

Received      Paper

**TOTAL:**

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

---

**(d) Manuscripts**

Received      Paper

**TOTAL:**

Number of Manuscripts:

Books

Received      Book

TOTAL:

Received      Book Chapter

TOTAL:

Patents Submitted

Patents Awarded

Awards

Graduate Students

<u>NAME</u>	<u>PERCENT_SUPPORTED</u>
FTE Equivalent:	
Total Number:	

Names of Post Doctorates

<u>NAME</u>	<u>PERCENT_SUPPORTED</u>
FTE Equivalent:	
Total Number:	

---

### Names of Faculty Supported

<u>NAME</u>	<u>PERCENT SUPPORTED</u>	National Academy Member
Lawrence Carin	0.00	
<b>FTE Equivalent:</b>	<b>0.00</b>	
<b>Total Number:</b>	<b>1</b>	

### Names of Under Graduate students supported

<u>NAME</u>	<u>PERCENT SUPPORTED</u>
<b>FTE Equivalent:</b>	
<b>Total Number:</b>	

### Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period: ..... 0.00

The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:..... 0.00

The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:..... 0.00

Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):..... 0.00

Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:..... 0.00

The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense ..... 0.00

The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields: ..... 0.00

---

### Names of Personnel receiving masters degrees

<u>NAME</u>
<b>Total Number:</b>

### Names of personnel receiving PHDs

<u>NAME</u>
<b>Total Number:</b>

### Names of other research staff

<u>NAME</u>	<u>PERCENT SUPPORTED</u>
<b>FTE Equivalent:</b>	
<b>Total Number:</b>	

---

Sub Contractors (DD882)

## **Inventions (DD882)**

### **Scientific Progress**

This workshop brought together leading experts in the new field of analysis and acquisition of high-dimensional data, including compressive sensing (CS). The meeting focused on new theory, algorithms and application. In addition to having talks from many of the leading researchers from academia, there were talks from the members of the US government and industry, on directions for this new technology, including future research directions and programs.

A full text report is attached.

### **Technology Transfer**

Not applicable.

## **Final Report**

Duke Workshop on High-Dimensional Data Sensing and Analysis  
Workshop Dates: July 26-28, 2011

Principal Investigator:

Lawrence Carin, Duke University



# Workshop Program

July 26-28, 2011

Location: Duke University

The video of all presentations and the associated talk slides are at  
[http://sahd.pratt.duke.edu/Videos\\_and\\_Slides.html](http://sahd.pratt.duke.edu/Videos_and_Slides.html)

This workshop brought together leading experts in the new field of analysis and acquisition of high-dimensional data, including compressive sensing (CS). The meeting focused on new theory, algorithms and application. In addition to having talks from many of the leading researchers from academia, there were talks from the members of the US government and industry, on directions for this new technology, including future research directions and programs.

## Abstracts

**Regularization with variance-mean mixtures, James Scott.** We provide a data-augmentation scheme that unifies many common regularized estimators into a single class. This leads to simple algorithms based on iterative least squares for fitting models involving arbitrary combinations of likelihood and penalty functions within the class. The class itself is quite large: for example, it includes quantile regression, support vector machines, logistic and multinomial logistic regression, topic models, and autologistic models, along with the usual ridge regression, lasso, bridge, and Student-t models. To arrive at this unified framework, we represent a wide class of objective functions as variance-mean mixtures of Gaussians involving both the likelihood and penalty functions. This generalizes existing theory based solely on variance mixtures for the penalty function, and allows the theory of conditionally normal linear models to be brought to bear on a much wider class of models. We focus on two possible choices of the mixing measures: the generalized inverse-Gaussian and Polya distributions, leading to the hyperbolic and Z distributions, respectively. We exploit this conditional normality to find sparse, regularized estimates using a new algorithm called tilted, iteratively re-weighted conjugate gradient. This algorithm is efficient, easily parallelizable, and offers a block update in situations where previous algorithms necessitate coordinate-wise updates. It also offers an exact MAP approach for estimating many models, including topic models, that have typically been fit using variational EM.

**A SURE Estimator for a Heteroscedastic Hierarchical Model, Lawrence D. Brown.** Hierarchical models are extensively studied and widely used in statistics and many other scientific areas. They provide an effective tool for combining information from similar resources and achieving partial pooling of inference. Since the seminal work by James and Stein (1961) and Stein (1962), shrinkage estimation has become one major focus for hierarchical models. For the homoscedastic normal model, it is well known that shrinkage estimators, especially the James-Stein estimator, have good risk properties. The heteroscedastic model, though more appropriate for practical applications, is less well

studied and it is unclear what types of shrinkage estimators are superior in terms of risk properties. We propose in this paper a class of shrinkage estimators based on the Stein's unbiased estimate of risk (SURE). We establish an asymptotic optimality property for these estimators as the number of means  $p \rightarrow \infty$ . We then extend our construction to create a class of semi-parametric shrinkage estimators and establish corresponding asymptotic optimality results. We emphasize that though the form of our SURE estimators is partially obtained through a normal-normal hierarchical model, their optimality properties do not heavily depend on such distribution assumptions. We apply the methods to two real data sets and obtain encouraging results. This is joint work with Samuel Kou and Xianchao Xie.

**The Return of PCA: Statistical Compressed Sensing of GMM and its Applications to Inverse problems, Guillermo Sapiro.**

A novel framework of compressed sensing, namely statistical compressed sensing (SCS), that aims at efficiently sampling a collection of signals that follow a statistical distribution, and achieving accurate reconstruction on average, is introduced. SCS based on Gaussian models is investigated in depth. For signals that follow a single Gaussian model, with Gaussian or Bernoulli sensing matrices of  $O(k)$  measurements, considerably smaller than the  $O(k \log(N/k))$  required by conventional CS based on sparse models, where  $N$  is the signal dimension, and with an optimal decoder implemented via linear filtering, significantly faster than the pursuit decoders applied in conventional CS, the error of SCS is shown tightly upper bounded by a constant times the best  $k$ -term approximation error, with overwhelming probability. The failure probability is also significantly smaller than that of conventional sparsity-oriented CS. Stronger yet simpler results further show that for any sensing matrix, the error of Gaussian SCS is upper bounded by a constant times the best  $k$ -term approximation with probability one, and the bound constant can be efficiently calculated. For Gaussian mixture models (GMMs), that assume multiple Gaussian distributions and that each signal follows one of them with an unknown index, a piecewise linear estimator is introduced to decode SCS. The accuracy of model selection, at the heart of the piecewise linear decoder, is analyzed in terms of the properties of the Gaussian distributions and the number of sensing measurements. A maximum a posteriori expectation-maximization algorithm that iteratively estimates the Gaussian models parameters, the signals model selection, and decodes the signals, is presented for GMM-based SCS. In real image sensing applications, GMM-based SCS is shown to lead to improved results compared to conventional CS, at a considerably lower computational cost. Joint work with Guoshen Yu and Stephane Mallat.

**Vector Diffusion Maps and the Connection Laplacian, Amit Singer.** Motivated by problems in structural biology, specifically cryo-electron microscopy, we introduce vector diffusion maps (VDM), a new mathematical framework for organizing and analyzing high dimensional data sets, images and shapes. VDM is a mathematical and algorithmic generalization of diffusion maps and other non-linear dimensionality reduction methods, such as LLE, ISOMAP and Laplacian eigenmaps. While existing methods are either directly or indirectly related to the heat kernel for functions over the data, VDM is based on the heat kernel for vector fields. VDM provides tools for organizing complex data sets, embedding them in a low dimensional space and interpolating and regressing vector fields over the data. In particular, it equips the data with a metric, which we refer to as the vector diffusion distance. In the manifold learning setup, where the data set is distributed on (or near) a low dimensional manifold  $M^d$  embedded in  $\mathbb{R}^p$ , we prove the relationship between VDM and the connection-Laplacian operator for vector fields over the manifold. Applications to structural biology (cryo-

electron microscopy and NMR spectroscopy), computer vision and shape space analysis will be discussed. Joint work with Hau-tieng Wu.

**Sparsity and Rank Minimization in Unions of Subspaces, Rene Vidal.** We consider the problem of fitting a union of subspaces to a collection of data points drawn from the subspaces and corrupted by noise/outliers. We pose this problem as a rank minimization problem, where the goal is to decompose the corrupted data matrix as the sum of a clean dictionary plus a matrix of noise/outliers. By constraining the dictionary to be self-expressive, i.e., its elements can be written as a linear combination of other elements, we formulate this problem as one of minimizing the nuclear norm of the matrix of linear combinations. Our first contribution is to show that, with noisy data, this problem can be solved in closed form from the SVD of the data matrix. Remarkably, this is true for both one or more subspaces. A key difference with respect to existing methods is that our framework results in a polynomial thresholding of the singular values with minimal shrinkage. Indeed, a particular case of our framework in the case of a single subspace leads to classical PCA, which requires no shrinkage. In the case of multiple subspaces, our framework provides an affinity matrix that can be used to cluster the data according to the subspaces. A major difference with respect to existing methods is that, rather than using the corrupted data as a dictionary, we build this affinity from a clean dictionary. Moreover, both the clean dictionary and the affinity matrix can be computed in closed form. In the case of data corrupted by outliers, a closed-form solution appears elusive. We thus use an augmented Lagrangian optimization framework, which requires a combination of our proposed polynomial thresholding operator with the more traditional shrinkage-thresholding operator.

**Alternating Direction Optimization: from Imaging Inverse Problems to Inference in Graphical Models, Mário A. T. Figueiredo.** This talk will cover our recent work on the application of a class of techniques known as ADMM (alternating direction method of multipliers, which belongs to the family of augmented Lagrangian methods) to several imaging inverse problems under non-smooth sparsity-inducing regularization, namely: standard image restoration/reconstruction from linear observations (e.g., compressive sensing) with a Gaussian noise model, using either analysis or synthesis regularization and unconstrained or constrained optimization formulations; image restoration from Poissonian observations, also using analysis or synthesis regularization. We show that, in all these cases, the proposed methods inherit the theoretic convergence guarantees of ADMM and achieve state-of-the-art performance in terms of speed. To further illustrate the flexibility of this class of methods, we show how it can be used to seamlessly and efficiently address other problems and formulations: hybrid analysis-synthesis regularization; group-sparsity regularization (with or without group overlap); maximum a posteriori (MAP) inference in probabilistic graphical models.

**The Convex Geometry of Linear Inverse Problems, Ben Recht.** Building on the success of generalizing compressed sensing to matrix completion, this talk discusses progress on further extending the catalog of objects and structures that can be recovered from partial information. I will focus on a suite of data analysis algorithms designed to decompose signals into sums of atomic signals from a simple---but not necessarily discrete---set. These algorithms are derived in a convex optimization framework that encompasses previous methods based on  $\ell_1$ -norm minimization and nuclear norm minimization for recovering sparse vectors and low-rank matrices. I will provide sharp estimates of the number of generic measurements required for exact and robust recovery of a variety

of structured models. I will then detail several example applications and describe how to scale the corresponding inference algorithms to massive data sets.

**Multitask Sparsity via Maximum Entropy Discrimination, Tony Jebara.** A multitask learning framework is developed for discriminative classification and regression where multiple large-margin classifiers are estimated for different prediction problems. A maximum entropy discrimination (MED) framework is used to derive the multitask algorithm which involves only convex optimization problems that are straightforward to implement. Four multitask scenarios are described. The first produces support vector machines that learn a shared sparse feature selection over the input space. The second method produces multiple support vector machines that learn a shared conic kernel combination. The third multitask method produces a pooled classifier as well as adaptively specialized individual classifiers. Finally, graphical model structure estimation is posed as a multitask sparse classification problem. The maximum entropy optimization problems are implemented via a sequential quadratic programming method which leverages recent progress in fast SVM solvers. Convergence bounds are provided by bounding the MED sparsifying cost function with a quadratic function and ensuring only a constant factor runtime increase above standard independent SVM solvers. Results are shown on multitask data sets and favor multitask learning over single-task or tabula rasa methods. Time permitting, I will also discuss results where, instead of maximizing margin, we maximize relative margin and obtain significant accuracy improvements.

**Matrix Sparse Coding, John Lafferty.** We study the problem of multivariate regression where the data are naturally grouped, and a regression or covariance matrix is to be estimated for each group. We propose an approach in which a large dictionary of low rank matrices is estimated across groups, and a sparse linear combination of the dictionary elements is estimated to form a model within each group. This exploits the same intuition behind sparse coding that has been successfully developed in computer vision. We propose algorithms for matrix sparse coding and show that the high-dimensional statistical convergence rate improves with the strength of our assumptions on the shared structure of the groups in the underlying model. Joint work with Min Xu (CMU).

**Principled Regularization for Matrix Factorization, Robert Bell.** Regularized matrix factorization has proven to be a very effective tool for dealing with high dimensional data, particularly in building recommender systems. L2 regularization is implemented by adding to the optimization criterion one of more terms that are proportional to the L2 norm of various sets of parameters. Implementation requires specifying the structure of these regularization terms and values for regularization parameters used as multipliers. We focus on the question of how many distinct regularization parameters should be used for a typical matrix factorization model. Theory and a simulation experiment suggest using more regularization parameters than are often used in practice. This is joint work with Suhrid Balakrishnan.

**The CLASH Operator, Volkan Cevher.** The least absolute shrinkage and selection operator (Lasso) for linear regression exploits the geometric interplay of the  $\ell_2$ -data error objective and the  $\ell_1$ -norm constraint to arbitrarily select sparse models. Guiding this uninformed selection process with sparsity models has been precisely the center of attention over the last decade in order to improve learning performance. To this end, we alter the selection process of Lasso in this talk to explicitly leverage combinatorial sparsity models (CSMs) via the combinatorial selection and least absolute shrinkage

operator (CLASH). A highlight is the introduction of a new algorithmic definition of CSMs, which we dub as the Polynomial time Modular epsilon-Approximation Property (PMAP\_epsilon). PMAP\_epsilon enables us to determine the impact of approximate combinatorial projections within CLASH. We then provide concrete guidelines how to leverage sets with PMAP\_epsilon within CLASH, and characterize CLASH's estimation guarantees as a function of epsilon as well as the set restricted isometry constants of the regression matrix. Finally, we present experimental results using both simulated and real world data to demonstrate the effectiveness of CLASH.

**Finding structure with randomness: Probabilistic algorithms for constructing low-rank matrix decompositions, Joel Tropp.** Computer scientists have long known that randomness can be used to improve the performance of algorithms. A familiar application is the process of dimension reduction, in which a random map transports data from a high-dimensional space to a lower-dimensional space while approximately preserving some geometric properties. By operating with the compact representation of the data, it is theoretically possible to produce approximate solutions to certain large problems very efficiently. Recently, it has been observed that dimension reduction has powerful applications in numerical linear algebra and numerical analysis. This talk provides a high-level introduction to randomized methods for computing standard matrix approximations, and it summarizes a new analysis that offers (nearly) optimal bounds on the performance of these methods. In practice, the techniques are so effective that they compete with—or even outperform—classical algorithms. Since matrix approximations play a ubiquitous role in areas ranging from information processing to scientific computing, it seems certain that randomized algorithms will eventually supplant the standard methods in some application domains. Joint work with Gunnar Martinsson and Nathan Halko. The paper is available at <http://users.cms.caltech.edu/~jtropp/papers/HMT11-Finding-Structure-SIREV.pdf>

**Fast First-Order and Alternating Direction Methods for Stable Principal Component Pursuit, Don Goldfarb** with Necdet Aybat and Garud Iyengar. Given a matrix  $M$  that is the sum of a low rank matrix  $X$  and a sparse matrix  $Y$ , it is known that under certain conditions,  $X$  and  $Y$  can be recovered with very high probability by solving a so-called robust principal component pursuit problem. When  $M$  has noise added to it, the recovery problem is called stable principal component pursuit. In this problem one minimizes the sum of the nuclear norm of  $X$  plus a scalar times the sum of the absolute values of the elements of  $Y$ , subject to the Frobenius norm of  $X+Y-M$  being less than some known noise level. We show how to apply Nesterov's optimal gradient and composite methods and a fast alternating linearization method to obtain epsilon-optimal solutions in  $O(1/\epsilon)$  iterations, while keeping the cost of each iteration low. We also propose a very efficient alternating direction augmented Lagrangian method, and prove its convergence.

**Operator-Theoretic Modeling for Radar in the Presence of Doppler, Doug Cochran.** Co-design of waveform and receiver processing is becoming increasingly important with the emergence of closed-loop waveform management, MIMO, and cooperating networked system concepts in radar. A common model of the radar scene in past work of this kind has been a linear time-invariant (LTI) operator with additive Gaussian noise that acts on each transmitted signal to produce the corresponding received signal. This model is intrinsically ill-suited to dynamic scenes or moving radar platforms because it cannot account for Doppler. This talk will discuss scene modelling based on Hilbert-Schmidt class (HS) operators on the space of finite-energy signals. This category of

models generalizes the LTI category in the sense that every LTI operator is also a HS operator, but the HS class includes operators that account for frequency shifts as well as time shifts, and are thus suitable for modelling radar scenes involving Doppler. Every HS operator is uniquely expressible as a superposition of elementary time and frequency shift operators, thus providing a convenient interpretation of a scene in terms of these physically meaningful operations on the transmitted signal. Application of this perspective to waveform design for target detection in noise and to optimal receiver processing for a given waveform to support target detection in clutter and noise are demonstrated. This is joint work with Stephen Howard and Bill Moran.

**Wavelets and filter banks on graphs, Pierre Vandergheynst.** In this talk I will first describe a simple construction of wavelets on graphs from spectral graph theory. Spectral graph wavelets are localized on the vertices of a (weighted) undirected graph and can be used to process data defined on networks for example. I will also describe an efficient algorithm to implement the transform and its adjoint which are the basic ingredients for general sparse recovery problems. Then I will describe our ongoing efforts to construct critically sampled filter banks generating wavelets on graphs. A neat connection with nodal theorems will allow us to introduce and study downsampling on graphs and a reduction technique will allow us to iterate the construction through scales. Interestingly, these techniques lead, for a class of graphs, to explicit perfect reconstruction equations that take the form of quadrature mirror conditions.

**Sequential Analysis in High-Dimensional Multiple Testing and Sparse Recovery, Robert Nowak.** The study of large-scale networked systems is now a major focus in engineering and science. Deciding what, where, when and how to sample such systems are crucial issues when measurement or experimental resources are limited. Standard experimental methods are non-adaptive: the experimental design is fixed before measurements are taken. One can envision alternative strategies in which the sampling or experimental design is optimized sequentially as more information is gained from previously collected data. In this talk, I will discuss and motivate the need for sequential, adaptive methods with a variety of applications. Then I will illustrate the main ideas by focusing on high-dimensional multiple hypothesis testing and sparse recovery problems, and their applications in biology, cognitive radio and compressed sensing. For example, consider testing to decide which of  $n > 1$  genes are differentially expressed in a certain disease. Suppose each test takes the form  $H_0: X \sim N(0,1)$  vs.  $H_1: X \sim N(m,1)$ , where  $N(m,1)$  is the Gaussian density with mean  $m$  and variance 1, and  $H_1$  represents the case of differential expression. The signal-to-noise ratio ( $SNR = m^2$ ) must be sufficiently large in order to guarantee a small probability of error. Non-sequential methods, which use an equal number of samples per test, require that the SNR grows like  $\log(n)$ . This is simply because the squared magnitude of the largest of  $n$  independent  $N(0,1)$  noises is on the order of  $\log(n)$ . Non-sequential methods cannot overcome this curse of dimensionality. Sequential testing, however, is capable of breaking the curse by adaptively focusing experimental resources on certain components at the expense of others. I will discuss a simple sequential method, in the spirit of classic sequential probability ratio testing that only requires the SNR to grow like  $\log(s)$ , where  $s$  is the number of tests in which the data are distributed according to  $H_1$ . In applications like gene testing,  $s$  is much much smaller than  $n$  and so the gains can be quite significant. For example, if  $s = \log n$ , then the sequential method is reliable if the SNR is  $O(\log \log n)$ , compared to the  $O(\log n)$  requirement of non-sequential approaches. The distinction is even more dramatic when the tests involve certain one-sided



distributions which arise, for example, in cognitive radio spectrum sensing. In this situation the gains can be doubly exponential in  $n$ .

**Applications and Implications of Fast Methods for L1 Related Optimization, Stan Osher.** New and newly revived fast methods for L1, total variation, tight frame based optimization have led to innumerable interesting applications. These include hyper spectral imaging, surveillance, learning theory, separation of dust from aerosols, retinex... I will discuss these methods, applications and future possibilities. This is joint work with many people.

**Sparse and Smooth: An optimal convex relaxation for high-dimensional kernel regression, Martin Wainwright.** The problem of non-parametric regression is well-known to suffer from a severe curse of dimensionality, in that the required sample size grows exponentially with the dimension  $d$ . Consequently, the success of any successful learning procedure in high dimensions must exploit some kind of low-dimensional structure. This talk focuses on non-parametric estimation within the family of sparse additive models, which consist of sums of univariate functions over  $s$  unknown co-ordinates. We derive a simple and intuitive convex relaxation for estimating sparse additive models described by reproducing kernel Hilbert spaces. The method involves solving a second-order cone program (SOCP), and so is suitable for large-scale problems. Working within a high-dimensional framework that allows both the dimension  $d$  and sparsity  $s$  to scale, we derive convergence rates that consist of two terms: a *subset selection term* that captures the difficulty of finding the unknown  $s$ -sized subset, and an *estimation error* that captures the difficulty of estimation over kernel classes. Using information-theoretic methods, we derive matching lower bounds on the minimax risk, showing that the SOCP-based method is optimal. Based on joint work with Garvesh Raskutti and Bin Yu, UC Berkeley. Arxiv paper: <http://arxiv.org/abs/1008.3654>

**Nonconvex Splitting Algorithms and Video Decomposition, Rick Chartrand.** There has been much recent work applying splitting algorithms to optimization problems designed to produce sparse solutions. In this talk, we'll look at extensions of these methods to the nonconvex case, motivated by results in compressive sensing showing that nonconvex optimization can recover signals from many fewer measurements than convex optimization. Our primary example will be the application of these methods to the decomposition of high-dimensional datasets, most notably video, into low-dimensional and sparse components.

**Recovering Simple Signals, Anna Gilbert.** The primary goal of compressed sensing and (non-adaptive) combinatorial group testing is to recover a sparse vector from an underdetermined set of linear equations. Both problems entail solving an undetermined linear system but they use different models of arithmetic, different models of randomness models for the matrix, and different guarantees upon the solution and the class of signals from which it is drawn. Lipton introduced a model for error correction where the channel is computationally bounded, subject to standard cryptographic assumptions, and produces the error vector  $x$  that must be found and then corrected. This has been extended to create more efficient schemes against polynomial and logspace bounded channels. Inspired by these results in error correction, we view compressed sensing and combinatorial group testing as an adversarial process, where Mallory the adversary produces the vector to be measured, with limited information about the matrix. We define a number of computationally bounded models for Mallory and show that there are significant gains (in the minimum number of measurements) to be

had by relaxing the model from adversarial to computationally or information-theoretically bounded, and not too much (in some cases, nothing at all) is lost by assuming these models over oblivious or statistical models. We also show that differences in adversarial power give rise to different lower bounds for the number of measurements required to defeat such an adversary. We use lower bounds in communication complexity to show upper bounds for recovery against a streaming adversary. We also give a characterization of when a bounded adversary can produce a “bad” vector for matrices with the Restricted Isometry Property (RIP). Joint work with Brett Hemenway, Atri Rudra, Martin Strauss, and Mary Wootters

**Parameter Estimation from Compressive Samples, Justin Romberg.** We will discuss recent results on embedding arbitrary sets using a random projection. These results relate the number of measurements required for a stable embedding to the geometry of the set. We will then apply these results to several parametric estimation problems, including frequency estimation and shift estimation using matched filtering.

**Scalable Topic Models, David Blei.** Probabilistic topic modeling provides an important suite of tools for the unsupervised analysis of large collections of documents. Topic modeling algorithms can uncover the underlying themes of a collection and decompose its documents according to those themes. This analysis can then be used for tasks like corpus exploration, document search, and a variety of prediction problems. Traditionally, topic models are fitted with “batch” algorithms, iteratively analyzing each document of the collection and then updating the parameters that describe the model. In this talk, I will describe a faster alternative based on online variational Bayes (VB). Online VB uses online stochastic optimization on the variational objective function with a natural gradient step. With online VB, we can handily analyze massive document collections, including those arriving in a stream. We demonstrate this approach by fitting a Bayesian nonparametric topic model to 3.3 million articles from Wikipedia. We demonstrate that online VB finds topic models as good or better than those found with batch VB, and in a fraction of the time. (Joint work with Matt Hoffman, Chong Wang and Francis Bach)

**Diffuse Interface Models on Graphs for Classification of High-Dimensional Data, Andrea Bertozzi.** There are currently several communities working on algorithms for classification of high dimensional data. I will discuss recent joint work with Arjuna Flenner that develops a class of variational algorithms combining recent ideas from spectral methods on graphs with nonlinear edge/region detection methods traditionally used in the PDE-based imaging community. The algorithms are based on the Ginzburg-Landau functional which has classical PDE connections to total variation minimization. Convex-splitting algorithms allow us to quickly find minimizers of the proposed model and take advantage of fast spectral solvers of linear graph-theoretic problems. I present diverse computational examples involving both basic clustering and semi-supervised learning for different applications. Case studies include feature identification in images, segmentation in social networks, and segmentation of shapes in high dimensional datasets.

**On a new Internet Traffic Matrix (Completion) Problem, Walter Willinger.** During the last 10 years, the problem of inferring the intra-domain traffic matrix of large networks (e.g., AT&T, Sprint) has morphed from an active research topic with important practical applications into a non-problem -- many of today's networks are now instrumented to measure the traffic matrix directly, without any



need for inference or estimation. However, it seems safe to predict that the next 10 years will see significant research activities on a new and more challenging traffic matrix problem; that is, inferring the Internet's inter-domain traffic matrix that describes the, say, hourly or daily traffic volumes exchanged between the various networks that define the Internet as a "network of networks." In this talk, I will discuss some of the challenges that arise when working with this matrix including the problem formulation itself, potential solution methods, and the all-important validation requirements.

**Easy Does It: Dense and Sparse Spectral Estimation Algorithms that Are User Parameter Free, Jian Li.** User parameter free algorithms, like the simple fast Fourier transform (FFT), are easy to use and are desirable for diverse practical applications. However, the data-independent FFT algorithm suffers from poor resolution and high sidelobe level problems. Many parametric, nonparametric, and sparse (semi-parametric) spectral estimation algorithms have been introduced in the literature. They possess super resolution and low sidelobe level properties. However, most of these algorithms are not user parameter free, making them difficult to use in practical applications. Many of these algorithms are sensitive to data model errors and/or require second-order statistics of the measurement vector. We present herein an iterative adaptive approach (IAA) for dense spectral estimation and a sparse learning via iterative minimization (SLIM) algorithm for sparse spectral estimation. Both algorithms are user parameter free, with the dense algorithm more accurate and the sparse algorithm computationally more efficient.

**Point process modeling for directed interaction networks, Patrick J. Wolfe.** Network data often take the form of repeated interactions between senders and receivers tabulated over time. We introduced a framework to treat these directed interactions as a multivariate point process: a Cox multiplicative intensity model using covariates that depend on the history of the process. We are able to show consistency and asymptotic normality for the resulting partial-likelihood-based estimators under suitable regularity conditions, and describe an efficient procedure for fitting. We also show how to explicitly treat multicast interactions--those involving a single sender but multiple receivers--in this framework. A motivating data example shows the effects of reciprocation and group-level preferences on message sending behavior in a corporate e-mail network. (Joint work with Patrick Perry)

**Sublinear Time, Measurement-Optimal, Sparse Recovery For All, Martin Strauss.** An approximate sparse recovery system in  $\ell_1$  norm formally consists of parameters  $N, k, \epsilon$  an  $m$ -by- $N$  measurement matrix,  $\Phi$ , and a decoding algorithm,  $D$ . Given a vector,  $x$ , where  $x_k$  denotes the optimal  $k$ -term approximation to  $x$ , the system approximates  $x$  by  $\hat{x} = D(\Phi x)$ , which must satisfy

$$\|\hat{x} - x\|_1 \leq (1 + \epsilon) \|x - x_k\|_1.$$

Among the goals in designing such systems are minimizing  $m$  and the runtime of  $D$ . We consider the "forall" model, in which a single matrix  $\Phi$  is used for all signals  $x$ . All previous algorithms that use the optimal number  $O(k \log(N/k))$  of measurements require superlinear time  $\Omega(N \log(N/k))$ . In this paper, we give the first algorithm for this problem that uses the optimum number of measurements (up to factors of  $\epsilon$ ) and runs in sublinear time  $o(N)$  when  $k = o(N)$ , assuming access to a data structure requiring space and preprocessing a bit more than  $N$ . Joint work with Ely Porat.

**Multi-task Averaging, Maya Gupta.** We investigate a similarity-based additive regularization for jointly estimating the means of  $T$  random variables from iid samples, which we refer to as multi-task averaging. We show that this simple multi-task formulation has surprising theoretical properties, and broad applicability to real problems.

**Smoothing Proximal Gradient Method for General Structured Sparse Regression, Eric Xing.** In many modern problems across areas such as genomics, computer vision, and natural language process, one is interested in learning a Sparse Structured Input-Output Regression Model (SIORM), in which the input variables of the model such as variations on a human genome bear rich structure due to the genetic and functional dependences between entities in the genome; and the output variables such as the disease traits are also structured because of their interrelatedness. A SIORM can nicely capture rich structural properties in the data, but raises severe computational and theoretical challenge on consistent model identification. In this talk, I will present algorithms and theories for learning Sparse SIORMs of various kinds in high dimensional input/output space, with a fast and highly scalable optimization procedure, and strong statistical guarantees. In particular, we investigate a smoothing proximal gradient method, which can solve structured sparse regression problems with a smooth convex loss and a wide spectrum of structured-sparsity-inducing penalties. Our method achieves a convergence rate significantly faster than the standard first-order method, subgradient method, and is much more scalable than the most widely used interior-point method. I will demonstrate application of our approach to problems in large-scale genome association analysis and web image understanding. This is joint work with Xi Chen and Seyoung Kim.

**Image acquisition using sparse (pseudo)-random matrices, Piotr Indyk.** I will present an overview of our work on sparse recovery/compressive sensing schemes that use \*sparse binary\* matrices. Such schemes provably offer good (near-optimal) compression rates and fast recovery times simultaneously. I will also discuss some of the existing imaging architectures that naturally utilize such matrices, as well as their applications.

**Shrinkage estimators for low rank matrices, Andrea Montanari.** Suppose to collect noisy observations of a low-rank or approximately low-rank matrix. A simple approach to estimating the matrix is to take its singular value decomposition retain a few singular values, and scale them appropriately. I will show that this approach is optimal (up to a universally multiplicative constant) even if a large part of the entries are missing, within a 'bounded entries' model. When studying the constant factor, I will show that substantial gain can be achieved by modifying the shrinking of singular values. I will use recent results in random matrix theory to evaluate various shrinkage transformations.

**Subspace Expanders and Low Rank Matrix Recovery, Babak Hassibi.** The adjacency matrix of an expander graph has been proven useful as the measurement matrix in sparse signal recovery since 1) it maps support sets in the ambient space to large (expanded) support sets in the measurement space (thereby allowing sparse recovery) and 2) it is a sparse matrix which allows for efficient (often linear time) recovery algorithms. In this talk, we generalize the notion of expansion from support sets to subspaces and prove the existence of subspace expanders, i.e., linear mappings from  $\mathbb{R}^{(n_1 \times n_1)}$  to  $\mathbb{R}^{(m_1 \times m_2)}$  that "expand" the rank of low rank matrices. We argue that such mappings are

suitable as the measurement operator for low rank matrix recovery problems and use them to develop fast recovery algorithms. We also mention several potential applications, such as quantum detection.

**Spectral classification sensors: An adaptive approach, Mike Gehm.** Frequently, the goal of spectroscopy or spectral imaging is not signal reconstruction, but rather classification of the input spectra against an existing spectral library. We have developed an adaptive approach that modifies the system measurement kernel in response to prior measurements and, as a result, achieves orders-of-magnitude faster time-to-classification than traditional methods. Design and optimization of the extremely high-dimensional measurement kernels is complicated by the need to simultaneously respect physical, system, and task constraints.

**An Efficient Dictionary for Reconstruction of Sampled Multiband Signals, Michael B. Wakin.** There remains a significant gap between the discrete, finite dimensional compressive sensing (CS) framework and the problem of acquiring a continuous-time signal. In this poster, we will discuss how sparse representations for multiband signals can be incorporated into the CS framework through the use of Discrete Prolate Spheroidal Sequences (DPSS's). DPSS's form a highly efficient basis for sampled bandlimited functions; by modulating and merging DPSS bases, one obtains a sparse representation for sampled multiband signals. We will discuss the use of DPSS bases for both signal recovery and the cancellation of strong narrowband interferers from compressive samples. Joint work with Mark A. Davenport.

## POSTERS

**Exploration & Representation of Data with Geometric Wavelets, Eric E Monson, Guangliang Chen, Rachael Brady & Mauro Maggioni.** Geometric Wavelets is a new multi-scale data representation technique which is useful for a variety of applications such as data compression, interpretation and anomaly detection. We have developed an interactive visualization with multiple linked views to help users quickly explore data sets and understand this novel construction. Currently the interface is being used by applied mathematicians to view results and gain new insights, speeding methods development.

**Sparse Modeling of Human Actions from Motion Imagery, Alexey Castrodad and Guillermo Sapiro.** An efficient sparse modeling pipeline for the classification of human actions from video is developed. Spatio-temporal features that characterize changes in the image are first extracted. This is followed by the learning of a class-structured dictionary encoding the individual actions of interest. Classification is then based on reconstruction, where the label assigned to each video comes from the optimal sparse linear combination of the learned basis vectors (action primitives) representing the learned actions. A low computational cost deep-layer model learning the interclass correlations of the data is added for increasing discriminative power. In spite of its simplicity and low computational cost, the method outperforms previously reported results for virtually all standard datasets.

**Stable Embeddings of Time Series Data, Han Lun Yap and Christopher J. Rozell.** Takens' theorem remarkably established that a delay coordinate map (DCM) can be a one-to-one mapping of a

low-dimensional attractor from the system state space. However, Takens' theorem is fragile as small imperfections can induce large errors in this attractor representation. Taking inspiration from compressed sensing, we extend Takens' result to establish deterministic, non-asymptotic sufficient conditions for a DCM to form a stable embedding in the restricted case of linear dynamical systems and observation functions. Using tools from stable manifold embeddings, we further establish that by filtering the DCM output, we could reduce the number of measurements without unduly sacrificing the conditioning of the resulting stable embedding.

**Concentration Inequalities and Isometry Properties for Compressive Block Diagonal Matrices, Han Lun Yap, Jae Young Park, Armin Eftekhari, Christopher Rozell, and Michael Wakin.** In this poster, we survey our recent analysis of randomized, compressive block diagonal matrices. We present concentration of measure bounds which indicate that (unlike dense i.i.d. random matrices) the probability of norm preservation actually depends on the signal being measured and discuss implications of this fact in various compressive signal processing applications. We also present an RIP bound for block diagonal matrices and explain that in the best case—for signals that are sparse in certain bases—these matrices perform nearly as well as dense i.i.d. random matrices despite having many fewer nonzero entries.

**Learning Sparse Codes for Hyperspectral Images, Adam Charles, Bruno Olshausen, and Christopher Rozell.** The exploitation of low dimensional structures in data has dramatically improved our ability to solve highly underdetermined linear inverse problems. While early results leverage sparsity in analytic dictionaries (such as wavelets) additional improvements have since been observed by adapting a dictionary to a specific data type. In this work we demonstrate the advantage of learning a spectral dictionary for Hyperspectral imagery (HSI). Specifically, we show that a dictionary learning procedure can learn relevant material spectra which can be used to infer materials in a HSI scene. The resulting codes can be used to super-resolve HSI data from lower resolution measurements (i.e., multispectral imagery), as well as to reduce the complexity of supervised classifiers for ground materials.

**Convergence Properties and Rate Analysis for Neural Networks for Sparse Approximation, Aurèle Balavoine, Justin Romberg, and Christopher Rozell.** We present an analysis of the Locally Competitive Algorithm (LCA), a Hopfield-style neural network that solves sparse approximation problems. This class of problems plays a significant role in both theories of neural coding and applications in signal processing, and may benefit from neural network approaches that could be implemented in fast, low power analog architectures. However, traditional network analysis approaches are difficult because the objective functions are non-smooth. Specifically, we characterize the convergence properties of this system by showing that the LCA is globally convergent to a fixed point corresponding to the exact solution of the objective function, and (under some mild conditions) this solution is reached in finite time. Furthermore, we characterize the convergence rate of the system by showing that the LCA converges exponentially fast with an analytically bounded rate (that depends on the specifics of a given problem). We support our analysis with several illustrative simulations. In the specific case of Compressed Sensing, where the dictionary usually has low coherence and the  $\ell_1$ -norm is used as the sparsity-inducing cost function, we show that it is possible to choose the threshold to achieve convergence in a small number of steps.

**A Compressive Sensing LIDAR Architecture, Darryl Sale, Christopher Rozell, Justin Romberg, and Aaron Lanterman.** In this work, we present a linear mode LIDAR sensor architecture based on the principles of Compressive Sensing. This conceptual architecture would produce high resolution range images without complicated scanning or detection electronics by exploiting the inherent sparsity of LIDAR reflections. Its complementary optical paths and linear mode detectors transform time-varying LIDAR photon arrivals into a compressed measurement waveform corresponding to each row of the sensing matrix. A small number of samples of these waveforms are used in standard sparse approximation algorithms to recover the scene in a format well-suited for parallel processing. MATLAB simulation results based on DIRSIG scene models demonstrate excellent detector performance with substantially fewer measurements than required by conventional scanning LIDAR.

**Principal Curvature Analysis: Dynamic Feature Selection By Moving Frames Optimization, Xiao Bian and Hamid Krim.** We propose a novel geometrical method to extract dynamic features of functional data in high dimensions based on moving frames optimization. Structural invariant dynamic features are represented by the corresponding curvature vectors. The objective functions are designed to optimize on moving frames in order to make the curvature vectors in a simple form, which we call principal curvature vectors. Their solutions are in closed form and have advantages to applied in pattern classification and recognition.

**High Dimensional Graphical Model Selection: Tractable Graph Families and Latent Trees, Anima Anandkumar.** Capturing complex interactions among a large set of variables is a challenging task. Probabilistic graphical models or Markov random fields provide a graph-based framework for capturing such dependencies. Graph estimation is an important task, since it reveals important relationships among the variables. I will present a unified view of graph estimation and propose a simple local algorithm for graph estimation using only low-order statistics of the data. We establish that the algorithm has consistent graph estimation with low sample complexity for a class of graphical models satisfying certain structural and parameter criteria. We explicitly characterize these model classes and point out interesting relationships between the graph structure and the parameter regimes, required for tractable learning. Many graph families such as the classical Erdos-Renyi random graphs, random regular graphs, and the small-world graphs can be learnt efficiently under our framework. The second part of the work is motivated by the following question: can we discover hidden influences acting on the observed variables? We consider latent tree models for capturing hidden relationships. We develop novel algorithms for learning the unknown high-dimensional latent tree structure. Our algorithm is amenable to efficient implementation of the Bayesian Information Criterion (BIC) to tradeoff the number of hidden variables with the accuracy of the model fitting. Experiment on the S&P 100 financial data reveals sectorization of the companies and experiment on the newsgroups data automatically categorizes words into different topics.

**Sparse Principal Components Analysis with Multiply Labeled Data, Kenneth D. Morton Jr., Leslie M. Collins, Peter A. Torriane.** Standard dimensionality reduction techniques such as principal components analysis (PCA) and factor analysis model high dimensional data using a small number of high dimensional basis vectors and a corresponding set of low dimensional scores. In general, the basis (or set of factors) is determined so as to maximize the variance in the set of scores, minimize the projection error, or maximize the likelihood of the assumed model. Unfortunately, these techniques ignore information contained in any available supplementary data related to the observations (e.g.,

observation labels). Although techniques such as linear discriminant analysis (LDA) and partial least squares discriminant analysis (PLSDA) can be used to infer a discriminative basis that incorporates observation labels, these techniques do not learn a generative basis while incorporating this additional information. Based on the Bayesian sparse factor analysis formulation of Chen et al. (2011) this work develops multi-label PCA (MLPCA), a generative dimension reduction technique that can incorporate supplementary information and an online variational Bayesian learning algorithm is developed for the parameters. The model automatically incorporates information from multiple labels for each observation, and is capable of learning a set of factors for each individual class even if data from a particular class is only observed “mixed” with data from different classes. By way of example, the MLPCA model is utilized within a detection algorithm for identifying multiple chemical residues on multiple substrates where identification of the factors corresponding to individual residues is important, and data from each residue is always measured in a mixture with a substrate. It is demonstrated that a detection algorithm utilizing MLPCA provides improved robustness to variations in substrate compared to standard chemometric techniques for residue detection.

**Joint Dynamic Sparse Representation for Classification, Haichao Zhang, Nasser M. Nasrabadi and Thomas S. Huang.** We propose a joint dynamic sparse representation algorithm for classification with multiple measurements /observations. We formulate classification task as a multi-input joint sparse representation model and exploit the correlations among the multiple inputs for classification using a novel joint dynamic sparsity prior. The proposed joint dynamic sparsity prior promotes shared joint sparsity pattern among the multiple sparse representation vectors at class-level, while allowing distinct sparsity patterns at atom-level within each class to facilitate a flexible representation. The proposed method can handle both homogenous as well as heterogeneous data within the same framework.

**Broadband Passive Spatial Spectrum Estimation in a Dynamic Environment, Jonathan Odom and Jeffrey Krolik.** Estimation of the time-varying acoustic field is essential for situational awareness in passive sonar. Adaptive processing often assumes both the field is stationary and the array is fixed for multiple observation windows. Highly dynamic scenarios such as high bearing rate sources or underwater maneuvers severely limit the utilization of multiple snapshots. Several models are considered for time-varying fields, and an online broadband maximum-likelihood estimator is introduced that is solved with an EM algorithm using as few as one snapshot. The number of estimated parameters can be reduced for broadband data when information, such as shape, about the source temporal spectra is known. Cramér-Rao bound analysis is used to understand effects of temporal spectrum knowledge on broadband processing. An example is given for the flat spectrum case to compare with conventional processing. The new spatial spectrum estimate leverages information across the temporal spectrum to reduce the impact of spatial grating lobes. Another feature of dynamic environments is array motion. Since underwater arrays are often subject to motion, the estimate must consider arbitrary, dynamic array shapes. Platforms such as autonomous underwater vehicles provide mobility but constrain the number of sensors. Exploiting maneuverable linear arrays with the new estimate allows for left-right disambiguation, more uniform angular resolution, and suppression of spatial grating lobes. Multi-source simulations are used to demonstrate the ability of a short, maneuvering array to reduce array backlobes as well as operate in the spatial grating lobe region. Detection performance of a weak high-bearing rate source in an interference dominated environment is evaluated for a flat spectrum.



**Nonparametric Multivariate Convex Regression with Applications to Value Function Approximation, Lauren Hannah and David Dunson.** We propose two new, nonparametric methods for multivariate regression subject to convexity or concavity constraints on the response function. Convexity constraints are common in economics, statistics, operations research and financial engineering, but there is currently no multivariate method that is computationally feasible for more than a few hundred observations. We introduce frequentist Convex Adaptive Partitioning (CAP) and Bayesian Multivariate Bayesian Convex Regression (MBCR), which both create a globally convex regression model from locally linear estimates fit on adaptively selected covariate partitions. Adaptive partitioning makes computation efficient even on large problems. We leverage the search procedure of CAP to create a computationally efficient reversible jump MCMC sampler for MBCR. We give strong consistency results for CAP in the univariate case and both weak and strong consistency results for MBCR in the general case. Convexity offers a few properties that can be exploited. First, it acts as a regularizer, making CAP and MBCR resistant to overfitting. Second, convex functions can be quickly minimized with commercial solvers. We use CAP and MBCR to fit value function approximation for real-world sequential decision problems with convex value-to-go functions. By allowing efficient search over the decision space, CAP and MBCR allow us much larger problems than can currently be solved with state of the art methods. We show that MBCR produces much more robust policies than CAP through model averaging.

**Efficient Gaussian Process Regression for Large Data Sets, Anjishnu Banerjee, David Dunson and Surya Tokdar.** Gaussian processes (GPs) are widely used in nonparametric regression, classification and spatio-temporal modeling, motivated in part by a rich literature on theoretical properties. However, a well known drawback of GPs that limits their use is the expensive computation, typically  $O(n^3)$  in performing the necessary matrix inversions with  $n$  denoting the number of data points. In large data sets, data storage and processing also lead to computational bottlenecks and numerical stability of the estimates and predicted values degrades with  $n$ . To address these problems, a rich variety of methods have been proposed, with recent options including predictive processes in spatial data analysis and subset of regressors in machine learning. The underlying idea in these approaches is to use a subset of the data, leading to questions of sensitivity to the subset and limitations in estimating fine scale structure in regions that are not well covered by the subset. Motivated by the literature on compressive sensing, we propose an alternative random projection of all the data points onto a lower-dimensional subspace. We demonstrate the superiority of this approach from a theoretical perspective and through the use of simulated and real data examples.

**Repulsive Mixtures, Francesca Petralia and David B. Dunson.** Mixture models have become extremely used for density estimation, clustering and as a component in flexible hierarchical models. In using mixture models for clustering, identifiability problems arise if mixture components are not sufficiently well separated and the data for the different sub-populations contain substantial overlap. Insufficiently separated components also creates problems in using mixture models for density estimation and robust modeling, as redundant components that are located close together can be introduced leading to an unnecessarily complex model as well as to various computational problems. Current practice in Bayesian mixture modeling generates the component-specific parameters from a common prior, which tends to favor components that are close together. As an alternative, we propose to generate mixture components from a repulsive process that favors placing components

further apart. Efficient algorithms are proposed for fully Bayes and fast approximate posterior computation allowing uncertainty in the number of components. The methods are illustrated using simulated data and several applications.

**Tensor factor models for high-dimensional unordered categorical data, Anirban Bhattacharya and David B. Dunson.** Gaussian latent factor models are routinely used for modeling of dependence in continuous, binary and ordered categorical data. For unordered categorical variables, Euclidean latent factor models lead to challenging computation and complex modeling structures. We relate the problem of modeling high-dimensional categorical variables to sparse tensor decomposition and propose a class of simplex factor models. Using a Bayesian approach for computation and inferences, an efficient MCMC algorithm is proposed that scales well with increasing dimensions. We provide a unified framework for testing associations and interactions among groups of variables and evaluate the approach through simulation examples and applications to modeling dependence in nucleotide sequences.

**Bayesian modeling of closed surfaces through tensor products, Debdeep Pati and David B. Dunson.** Closed surfaces provide a useful model for 3d shapes, with the data typically consisting of a cloud of points in  $\mathbb{R}^3$ . The existing literature on closed surface modeling focuses on frequentist point estimation methods that join surface patches along the edges, with surface patches created via Bezier surfaces or tensor products of B-splines. However, the resulting surfaces are not smooth along the edges and the geometric constraints required to join the surface patches lead to computational drawbacks. In this article, we develop a Bayesian model for closed surfaces based on tensor products of a cyclic basis resulting in infinitely smooth surface realizations. We impose sparsity on the control points through a double-shrinkage prior. Theoretical properties of the support of our proposed prior are studied and it is shown that the posterior achieves the optimal rate of convergence under reasonable assumptions on the prior. Several simulation examples are considered and the methods are applied to real data. We also extend our model for a single closed surface to a hierarchical model to accommodate multiple closed surfaces.

**Bayesian Kernel Mixtures for Counts, Antonio Canale and David B. Dunson.** Although Bayesian nonparametric mixture models for continuous data are well developed, there is a limited literature on related approaches for count data. A common strategy is to use a mixture of Poissons, which unfortunately is quite restrictive in not accounting for distributions having variance less than the mean. Other approaches include mixing multinomials, which requires finite support, and using a Dirichlet process prior with a Poisson base measure, which does not allow smooth deviations from the Poisson. As a broad class of alternative models, we propose to use nonparametric mixtures of rounded continuous kernels. An efficient Gibbs sampler is developed for posterior computation, and a simulation study is performed to assess performance. Focusing on the rounded Gaussian case, we generalize the modeling framework to account for multivariate count data, joint modeling with continuous and categorical variables, and other complications. The methods are illustrated through applications to marketing data.

**Bayesian Nonparametric Covariance Regression, Emily Fox and David B. Dunson.** Although there is a rich literature on methods for allowing the variance in a univariate regression model to vary with predictors, time and other factors, relatively little has been done in the multivariate case. Our



focus is on developing a class of nonparametric covariance regression models, which allow an unknown  $p \times p$  covariance matrix to change flexibly with predictors. The proposed modeling framework induces a prior on a collection of covariance matrices indexed by predictors through priors for predictor-dependent loadings matrices in a factor model. In particular, the predictor-dependent loadings are characterized as a sparse combination of a collection of unknown dictionary functions (e.g., Gaussian process random functions). The induced covariance is then a regularized quadratic function of these dictionary elements. Our proposed framework leads to a highly-flexible, but computationally tractable formulation with simple conjugate posterior updates that can readily handle missing data. Theoretical properties are discussed and the methods are illustrated through simulations studies and an application to the Google Flu Trends data.

**Bayes variable selection in semiparametric linear models, Suprateek Kundu and David B. Dunson.** There is a rich literature proposing methods and establishing asymptotic properties of Bayesian variable selection methods for parametric models, with a particular focus on the normal linear regression model and an increasing emphasis on settings in which the number of candidate predictors ( $p$ ) diverges with sample size ( $n$ ). Our focus is on generalizing methods and asymptotic theory established for mixtures of  $g$ -priors to semiparametric linear regression models having unknown residual densities. Using a Dirichlet process location mixture for the residual density, we propose a semiparametric  $g$ -prior which incorporates an unknown matrix of cluster allocation indicators. For this class of priors, posterior computation can proceed via a straightforward stochastic search variable selection algorithm. In addition, Bayes factor consistency is shown to result under various cases including proper priors on  $g$  and  $p$  diverging at a rate less than  $n$ .

**Flexible Scale Mixtures of Normals for Shrinkage Estimation, Artin Armagan, David Dunson and Merlise Clyde.** In recent years, a rich variety of shrinkage priors have been proposed that have great promise in addressing massive regression problems. In general, these new priors can be expressed as scale mixtures of normals, but have more complex forms and better properties than traditional Cauchy and double exponential priors. We first propose a new class of normal scale mixtures through a novel generalized beta distribution that encompasses many interesting priors as special cases. This encompassing framework should prove useful in comparing competing priors, considering properties and revealing close connections. We then develop a class of variational Bayes approximations through the new hierarchy presented that will scale more efficiently to the types of truly massive data sets that are now encountered routinely.

**An Empirical-Bayes Approach to Compressive Sensing via Approximate Message Passing, Jeremy Villa.** Traditional approaches to compressive sensing (e.g., Lasso, basis-pursuit denoising) take a worst-case approach to signal recovery, and are able to give guarantees for the recovery of any deterministic signal. Furthermore, recent work has connected these traditional approaches to minimax optimal estimation. Although the minimax property is admirable from a robustness perspective, estimation performance can be improved when the signal prior is known to differ from the least-favorable one. At the same time, care must be taken not to assume too much about the signal prior. In this work, we propose an empirical Bayesian approach to compressive sensing, where we assume an i.i.d Gaussian-Mixture signal model and an i.i.d Gaussian noise model, but make no assumptions on the parameters of the signal distribution and the noise distribution. We then employ the EM algorithm to jointly learn the distributional parameters (in an ML sense) and recover the signal (in an MMSE

sense), leveraging the recently proposed approximate message passing (AMP) algorithm for compressive sensing.

**Two Proposals for Robust PCA using Semidefinite Programming, Michael McCoy and Joel A. Tropp.** The performance of principal component analysis (PCA) suffers badly in the presence of outliers. This poster outlines two novel approaches for robust PCA based on programming. The first method, maximum mean absolute deviation rounding (MDR), seeks directions of large spread in the data while damping the effect of outliers. The second method produces a low-leverage decomposition (LLD) of the data that attempts to form a low-rank model for the data by separating out corrupted observations. This work also presents efficient computational methods for solving these SDPs. Numerical experiments confirm the value of these new techniques."

**Fast level set estimation from projection measurements, Kalyani Krishnamurthy, Waheed Bajwa, Rebecca Willett and Robert Calderbank.** Estimation of the level set of a function (i.e., regions where the function exceeds some value) is an important problem with applications in digital elevation maps, medical imaging, and astronomy. In many applications, however, the function of interest is acquired through indirect measurements, such as tomographic projections, coded-aperture measurements, or pseudo-random projections associated with compressed sensing. This poster describes a new methodology and associated theoretical analysis for rapid and accurate estimation of the level set from such projection measurements. The proposed method estimates the level set from projection measurements without an intermediate function reconstruction step, thereby leading to significantly faster computation. In addition, the coherence of the projection operator and McDiarmid's inequality are used to characterize the estimator's performance.

**Estimation of Dynamic Social Network Structure via Online Convex Programming, Maxim Raginsky, Corinne Horn, and Rebecca Willett.** Consider a dynamic social network composed of  $p$  agents, where the designation "dynamic" means that the network topology or, more broadly, the influence of each agent on the other agents, may evolve with time. We collect sequential observations of the agents' actions, and would like to use them to infer something in real time about the structure of the network. From the modeling point of view, both the topology of the network and the strengths of the agents' influence can be encoded in a time-varying sparse binary pairwise Markov random field (also known in statistical physics as the Ising model). Thus, our goal is to infer the parameters of such a model given the observed binary co-occurrences representing the agents' actions. The offline (batch) version of this problem was recently treated by Banerjee et al. [1], Ravikumar et al. [2], Kolar and Xing [3], and Hofling and Tibshirani [4]. As noted in [1] and [2], sparsity regularization can play a critical role in accurately estimating network structure from a relatively limited amount of data; this effect has been noted both empirically and theoretically in an analysis of sample complexity of different methods. By contrast, we work in the online setting, i.e., when the observations are available sequentially, one at a time. Moreover, even if all the observations are available at once, online algorithms often offer a computationally feasible and robust alternative to batch inference [5]. A notable feature of our approach is that we do not assume that the dynamical evolution of the agents' actions is, in fact, described by a time-varying Ising model; nor do we assume that sequential observations are conditionally independent. Rather, we treat the Ising model as a descriptive tool, which allows for easy interpretation of the observed network behavior in terms of a time-varying weighted graph, where the presence of a nonzero weight on an edge connecting a particular pair of

agents indicates mutual influence, while the sign of the weight indicates the nature of the correlation between the agents' actions (positive or negative). Our theoretical results show that we can use the sequential observations of the agents' actions to construct a sequence of network structure estimates that is nearly as good as if all the observations had been available at once. Finally, we will demonstrate the efficacy of the proposed approach on a simulated data set generated from a time-varying binary Markov random field, as well as on the US Senate voting records.

**Fast Imaging with Slow Cameras, Dikpal Reddy, Ashok Veeraraghavan, Ramesh Raskar, Rama Chellappa.** Over the years, the spatial resolution of cameras has steadily increased but the temporal resolution has remained the same. In this talk, I will present my work on converting a regular slow camera into a faster one. We capture and accurately reconstruct fast events using our slower prototype camera by exploiting the temporal redundancy in videos. First, I will show how by fluttering the shutter during the exposure duration of a slow 25fps camera we can capture and reconstruct a fast periodic video at 2000fps. Next, I will present its generalization where we show that per-pixel modulation during exposure, in combination with brightness constancy constraints allows us to capture a broad class of motions at 200fps using a 25fps camera. In both these techniques we borrow ideas from compressive sensing theory for acquisition and recovery.

**The Hierarchical Beta Process for Convolutional Factor Analysis and Deep Learning, Bo Chen, Gungor Polatkan, Guillermo Sapiro, David B. Dunson, Lawrence Carin.** A convolutional factor-analysis model is developed, with the number of filters (factors) inferred via the beta process (BP) and hierarchical BP, for single-task and multi-task learning, respectively. The computation of the model parameters is implemented within a Bayesian setting, employing Gibbs sampling; we explicitly exploit the convolutional nature of the expansion to accelerate computations. The model is used in a multi-level ("deep") analysis of general data, with specific results presented for image-processing data sets, e.g., classification.

**Tree-Structured Infinite Sparse Factor Model, XianXing Zhang, David B. Dunson, Lawrence Carin.** A tree-structured multiplicative gamma process (TMGP) is developed, for inferring the depth of a tree-based factor-analysis model. This new model is coupled with the nested Chinese restaurant process, to nonparametrically infer the depth and width (structure) of the tree. In addition to developing the model, theoretical properties of the TMGP are addressed, and a novel MCMC sampler is developed. The structure of the inferred tree is used to learn relationships between high-dimensional data, and the model is also applied to compressive sensing and interpolation of incomplete images.

**Dependent Hierarchical Beta Process for Image Interpolation and Denoising, Mingyaun Zhou, Hongxia Yang, Guillermo Sapiro, David Dunson and Lawrence Carin.** A dependent hierarchical beta process (dHBP) is developed as a prior for data that may be represented in terms of a sparse set of latent features, with covariate-dependent feature usage. The dHBP is applicable to general covariates and data models, imposing that signals with similar covariates are likely to be manifested in terms of similar features. Coupling the dHBP with the Bernoulli process, and upon marginalizing out the dHBP, the model may be interpreted as a covariate-dependent hierarchical Indian buffet process. As applications, we consider interpolation and denoising of an image, with covariates defined by the location of image patches within an image. Two types of noise models are considered: (i) typical white Gaussian noise; and (ii) spiky noise of arbitrary amplitude, distributed uniformly at random. In

these examples, the features correspond to the atoms of a dictionary, learned based upon the data under test (without a priori training data). State-of-the-art performance is demonstrated, with efficient inference using hybrid Gibbs, Metropolis-Hastings and slice sampling.

**On the Integration of Topic Modeling and Dictionary Learning, Lingbo Li, Guillermo Sapiro and Lawrence Carin.** A new nonparametric Bayesian model is developed to integrate dictionary learning and topic model into a unified framework. The model is employed to analyze partially annotated images, with the dictionary learning performed directly on image patches. Efficient inference is performed with a Gibbs-slice sampler, and encouraging results are reported on widely used datasets.

**Nonparametric Bayesian Modeling of Heterogeneous Space-Time Information Sources, Priyadip Ray, David Dunson and Lawrence Carin.** We propose a nonparametric Gaussian process (GP) factor analysis model for space-time-varying data. The problem of spatial inhomogeneity is addressed via a novel kernel stick-breaking process (KSBP) based mixture of GPs. We also propose a joint factor analysis technique to simultaneously model multi-modal correlated spatio-temporal information sources. We discuss theoretical properties of the KSBP-GP factor model and develop an MCMC algorithm for posterior inference. The performance of the proposed approach is demonstrated on the analysis of multi-year unemployment rates at various geographical locations, with these data analyzed jointly with time-evolving stock prices of companies in the S & P 500.

**Time-Evolving Modeling of Social Networks, Eric Wang, Jorge Silva, Rebecca Willett and Lawrence Carin.** A statistical framework for modeling and prediction of binary matrices is presented. The method is applied to social network analysis, specifically the database of US Supreme Court rulings. It is shown that the ruling behavior of Supreme Court judges can be accurately modeled by using a small number of latent features whose values evolve with time. The learned model facilitates the discovery of inter-relationships between judges and of the gradual evolution of their stances over time. In addition, the analysis in this paper extends previous results by considering automatic estimation of the number of latent features and other model parameters, based on a nonparametric-Bayesian approach. Inference is efficiently performed using Gibbs sampling.

**Audio source modeling based on structured sparsity and GMM using temporal constraints, Pablo Sprechmann.** In this work we present a new model for representing single tone audio sources. This includes single pitch musical instruments as well as human voice. Dictionary learning and Non negative matrix factorization (as well as its probabilistic counterparts) have been widely used for performing source separation. For each sound source, these algorithms learn a dictionary of spectral vectors to best represent it. These approaches have shown very good results in a variety of situations. However, when handling single tone audio sources, there is no guarantee that the coded signal will have a single tone. Additionally, no temporal structure is usually used in the coding, which is a very important aspect of audio sources. We present a simple model that is based on structured sparsity and Gaussian Mixture Models (GMM). For each source, we jointly learn a GMM for representing the spectral representation as well as a Markov chain that describes the structure of changes between these gaussians. We interpret the source coding as an inverse problem adding additional harmonic constraints.

**Non-Local Methods with Shape-Adaptive Patches (NLM-SAP), Charles Deledalle, Vincent Duval, Joseph Salmon.** We propose an extension of the Non-Local Means (NL-Means) denoising algorithm. The idea is to replace the usual square patches used to compare pixel neighborhoods with various shapes that can take advantage of the local geometry of the image. We provide a fast algorithm to compute the NL-Means with arbitrary shapes thanks to the Fast Fourier Transform. We then consider local combinations of the estimators associated with various shapes by using Stein's Unbiased Risk Estimate (SURE). Experimental results show that this algorithm improves the standard NL-Means performance and is close to state-of-the-art methods, both in terms of visual quality and numerical results. Moreover, common visual artifacts usually observed by denoising with NL-Means are reduced or suppressed thanks to our approach.

**Application of compressive sensing methods to image reconstruction in X-ray based tomography, Emil Y. Sidky and Xiaochuan Pan.** The compressive sensing idea of exploiting gradient sparsity for reducing scanning effort in computed tomography (CT) has proved to be quite fruitful. A survey of results, by us and other researchers, obtained with actual CT scanner data will be shown. These results include reconstruction from sparse-view data, high resolution imaging with low intensity X-ray illumination, and time-resolved imaging. As there is no theoretical result for linking image-gradient sparsity and CT sampling conditions yet, we will discuss our preliminary attempts at making this link with computer simulations.

**Sparse Measurement Matrices for Sequential Adaptive Compressive Sensing, Jarvis Haupt, Richard Baraniuk, Rui Castro, and Robert Nowak.** The theory of Distilled Sensing (DS) establishes that sequential adaptive sensing provides significant quantifiable performance improvements over non-adaptive sensing in noisy high dimensional testing problems. The DS idea was recently extended to highly undersampled settings typical in compressed sensing (CS), and it was shown that a Compressive Distilled Sensing (CDS) method can similarly result in performance gains in noisy CS problems. However, the results obtained in the initial work on CDS were restricted to settings where the unknown sparse signals of interest have limited dynamic range (on the order of a constant). In this work we describe a new adaptive compressive sensing approach that utilizes sequentially designed sparse measurement matrices. Our analysis of this new approach shows that similar performance improvements in adaptive compressive sensing can be achieved, but without restrictive assumptions on the dynamic range of the signal being acquired.

**Multiscale Analysis of Planar Arrangements, Mauro Maggioni and Guangliang Chen.** Modeling high-dimensional data by multiple low-dimensional planes is an important problem in many applications such as computer vision and pattern recognition. In the most general setting where only coordinates of the data are given, the problem asks to determine the optimal model parameters (i.e. number of planes and their dimensions), estimate the model planes, and cluster the data accordingly. Though many algorithms have been proposed, most of them need to assume prior knowledge of the model parameters and thus address only the last two components of the problem. In this paper we propose an efficient algorithm based on multiscale SVD analysis and spectral methods to tackle the problem in full generality. We also demonstrate its state-of-the-art performance on both synthetic and real data, in tasks such as motion segmentation and a simple scenario for face recognition.

**Compressive Sensing for Low Complexity Video Encoding, M. Salman Asif, Felix Fernandes, and Justin Romberg.** With the proliferation of inexpensive camcorders and cellphones, User-Generated-Content (UGC) has become commonplace and there is a need for Low Complexity Video Encoding (LoCVE) technology that can be deployed in these low-cost devices. Compressive Sensing (CS) is a new signal acquisition framework which enables the reconstruction of sparse signals from a small number of linear measurements. To reduce video-encoder complexity, we present a CS-based encoding scheme. Modern video-encoder complexity arises mainly from the transform-coding and motion-estimation blocks. In our proposed LoCVE scheme, we eliminate these blocks entirely: the encoder achieves compression by merely taking a few linear measurements of each video frame using modified imaging hardware. To guarantee stable reconstruction of the original video frames from only a few measurements, the decoder must effectively exploit the spatial and temporal redundancy in the underlying video frames. To leverage these redundancies, we use a coarse-to-fine reconstruction algorithm that iteratively refines motion estimates within adjacent frames, and recovers sparse wavelet coefficients. Critically-sampled wavelet transforms are shift-variant and cannot be used for transform-domain motion estimation. Therefore, we employ a near shift-invariant wavelet transform to estimate inter-frame motion, to incorporate statistical models that guide the recovery process and to reduce computational complexity.

**The Value of Redundant Measurement in Compressed Sensing, Victoria Kostina, Marco F. Duarte, Sina Jafarpour, and Robert Calderbank.** The aim of compressed sensing is to recover attributes of sparse signals using very few measurements. Given an overall bit budget for quantization, this paper demonstrates that there is value to redundant measurement. The measurement matrices considered here are required to have the property that signal recovery is still possible even after dropping certain subsets of  $D$  measurements. It introduces the concept of a measurement matrix that is weakly democratic in the sense that the amount of information about the signal carried by each of the designated  $D$ -subsets is the same. Examples of deterministic measurement matrices that are weakly democratic are constructed by exponentiating codewords from the binary second order Reed Muller code. The value in rejecting  $D$  measurements that are on average larger, is to be able to provide a finer grid for vector quantization of the remaining measurements, even after discounting the original budget by the bits used to identify the reject set. Simulation results demonstrate that redundancy improves recovery SNR, sometimes by a wide margin. Optimum performance occurs when a significant fraction of measurements are rejected.

**Gigapixel Binary Sensing: Image Acquisition Using Oversampled One-Bit Poisson Statistics, Yue M. Lu.** Before the advent of digital image sensors, photography, for the most part of its history, used film to record light information. In this work, we propose a new digital image sensor that is reminiscent of photographic film. Each pixel in the sensor has a binary response, giving only a one-bit quantized measurement of the local light intensity. To analyze its performance, we formulate the binary sensing scheme as a parameter estimation problem based on quantized Poisson statistics. We show that, with a single-photon quantization threshold and large oversampling factors, the Cramer-Rao lower bound of the estimation variance approaches that of an ideal unquantized sensor, that is, as if there were no quantization in the sensor measurements. Furthermore, this theoretical performance bound is shown to be asymptotically achievable by practical image reconstruction algorithms based on maximum likelihood estimators. Numerical results on both synthetic data and images taken by a prototype sensor verify the theoretical analysis and the effectiveness of the proposed image

reconstruction algorithm. They also demonstrate the benefit of using the new binary sensor in applications involving high dynamic range imaging. Joint work with Feng Yang, Luciano Sbaiz and Martin Vetterli. arXiv:1106.6954.v1 [[cs.IT](#)]